



THE UNIVERSITY *of* EDINBURGH

Edinburgh Research Explorer

## Weighted Low-Rank Approximation of Matrices and Background Modeling

### Citation for published version:

Dutta, A, Li, X & Richtarik, P 2018 'Weighted Low-Rank Approximation of Matrices and Background Modeling' ArXiv.

### Link:

[Link to publication record in Edinburgh Research Explorer](#)

### General rights

Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

### Take down policy

The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact [openaccess@ed.ac.uk](mailto:openaccess@ed.ac.uk) providing details, and we will remove access to the work immediately and investigate your claim.



# Weighted Low-Rank Approximation of Matrices and Background Modeling

Aritra Dutta, Xin Li, and Peter Richtárik

**Abstract**—We primarily study a special a weighted low-rank approximation of matrices and then apply it to solve the background modeling problem. We propose two algorithms for this purpose: one operates in the batch mode on the entire data and the other one operates in the batch-incremental mode on the data and naturally captures more background variations and computationally more effective. Moreover, we propose a robust technique that learns the background frame indices from the data and does not require any training frames. We demonstrate through extensive experiments that by inserting a simple weight in the Frobenius norm, it can be made robust to the outliers similar to the  $\ell_1$  norm. Our methods match or outperform several state-of-the-art online and batch background modeling methods in virtually all quantitative and qualitative measures.

**Index Terms**—Weighted low-rank approximation,  $\ell_1$ -norm minimization, Robust PCA, background modelling.

## I. INTRODUCTION

WE give a brief review of the classical low rank approximation of matrices and introduce the background modeling problem.

### A. Low-rank approximation

The standard low rank approximation aka the principal component analysis (PCA) problem can be defined as an approximation to a given matrix  $A \in \mathbb{R}^{m \times n}$  by a rank  $r$  matrix under the Frobenius norm:

$$X^* = \arg \min_{\substack{X \in \mathbb{R}^{m \times n} \\ \text{rank}(X) \leq r}} \|A - X\|_F^2, \quad (1)$$

where  $\|\cdot\|_F$  denotes the Frobenius norm of matrices. The solutions to (1) are given by

$$X^* = H_r(A) := U \Sigma_r V^T, \quad (2)$$

where  $A$  has singular value decompositions  $A = U \Sigma V^T$ , and  $\Sigma_r(A)$  is the diagonal matrix obtained from  $\Sigma$  by hard-thresholding operation that keeps only  $r$  largest singular values and replaces the other singular values by 0 along the diagonal. This is also referred to as Eckart-Young-Mirsky's theorem ([1]) and is closely related to the PCA method in statistics [2]. In image processing, rank-reduced signal processing, computer vision, and in many other engineering applications

SVD is a successful dimension reduction tool. The low rank matrix obtained through PCA is a good approximation to the data matrix  $A$  if  $A$  contains only normally (and independently) distributed noise. But, in many real world problems, if sparse large errors or outliers are present in the data matrix, PCA fails to deal with it and thus additional regularization has been introduced to accommodate the sparse outliers. One can think the background modeling problem in video sequences as a good example of such a real world problem. In the next section we will review the classic background modeling problem in light of matrix decomposition and briefly survey several state-of-the-art algorithms used to solve it.

### B. Background modeling in matrix decomposition framework

Background modeling and moving object detection are two key steps in many computer vision systems and video-surveillance applications. In the past decade, one of the most prevalent approaches used in background estimation is to treat it as a matrix decomposition problem ([3], [4], [5]). Given a sequence of  $n$  video frames with each frame converted into a vector  $\mathbf{a}_i \in \mathbb{R}^m$ ,  $i = 1, 2, \dots, n$ , the data matrix  $A = (\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n) \in \mathbb{R}^{m \times n}$  is the collection of all the frame vectors. Therefore, it is natural to consider a matrix decomposition problem by decomposing  $A$  as the sum of its background and foreground:

$$A = B + F,$$

where  $B, F \in \mathbb{R}^{m \times n}$  are the background and foreground matrices, respectively. The above problem is ill-posed, and it requires more information about the structure of the decomposition. In practice, the background  $B$  is expected to stay static or close to static throughout the frames when the camera motion is small and so  $B$  is assumed to be low-rank [6]. At the same time, the foreground,  $F$  is usually sparse if its size is relatively small compared to the frame size ([3], [4], [7], [8], [9]). These and similar observations leads to the models of the form ([3], [4], [8], [9], [10], [11], [12], [13], [14], [15]):

$$\min_{\substack{B, F \\ A=B+F}} f_{\text{rank}}(B) + f_{\text{sparse}}(F), \quad (3)$$

where  $f_{\text{rank}}$  is a function that encourages the rank of  $B$  to be low, and  $f_{\text{sparse}}$  is a function that encourages the foreground  $F$  to be sparse. Next we will discuss how this idea transformed into several state-of-the-art algorithms to solve the background modeling problem.

a) *Robust principal component analysis (RPCA)*: By using the above idea, RPCA [7], [9], [8] was introduced to solve the background modeling problem by considering the background frames,  $B$ , having a low-rank structure and the

Aritra Dutta is with the Visual Computing Center, Division of Computer, Electrical and Mathematical Sciences and Engineering (CEMSE) at King Abdullah University of Science and Technology, Thuwal, Saudi Arabia-23955-6900, e-mail: aritra.dutta@kaust.edu.sa (see <https://aritraddutta.weebly.com/>)

Xin Li is with the Department of Mathematics, University of Central Florida, FL, USA-32816, email: xin.li@ucf.edu.

Peter Richtárik is with the Visual Computing Center, Division of Computer, Electrical and Mathematical Sciences and Engineering (CEMSE) at King Abdullah University of Science and Technology, University of Edinburgh, and MIPT, e-mail: peter.richtarik@kaust.edu.sa (see <http://www.maths.ed.ac.uk/~prichter/>).

foreground,  $A - B$ , sparse. The nuclear norm (sum of the singular values) is used on background matrix  $B$  as a surrogate of rank and the  $\ell_1$  norm is used to encourage sparsity in the foreground:

$$\min_B \|A - B\|_{\ell_1} + \lambda \|B\|_*, \quad (4)$$

where  $\|\cdot\|_{\ell_1}$  and  $\|\cdot\|_*$  denote the  $\ell_1$  norm and the nuclear norm of matrices, respectively. RPCA is considered to be one of the most widely used state-of-art approaches to solve the background modeling problem. Several algorithms have been proposed to solve RPCA [7], [9], [8], for example, inexact and exact augmented Lagrange method (iEALM and EALM), accelerated proximal gradient (APG), just to name a few. However, RPCA cannot take advantage of any possible extra prior information on the background and can not be used as a supervised learning method.

*b) Generalized fused Lasso (GFL):* Consider a supervised learning situation when some pure background frames are given. Recently in [16], Xin et. al. proposed a supervised model called the GFL to address these issues. Let the data matrix  $A$  be written into  $A = (A_1 \ A_2)$ , where  $A_1$  contains the given pure background frames. Xin et. al. [16] proposed the following model: Find the background matrix  $B = (B_1 \ B_2)$ , and foreground matrix  $F = (F_1 \ F_2)$ , partitioned in the same way as  $A$ , such that

$$\min_{\substack{B, F \\ B_1 = A_1 \\ A_2 = B_2 + F_2}} \text{rank}(B) + \|F\|_{GFL}, \quad (5)$$

where  $\|\cdot\|_{GFL}$  denotes a norm that combines the  $\ell_1$  norm and a local spatial total variation norm (to encourage connectivity of the foreground). To make the problem more tractable, Xin et. al. further specialized the above model by assuming  $\text{rank}(B) = \text{rank}(B_1)$ . Since  $B_1 = A_1$  and  $A_1$  is given, so  $r := \text{rank}(B_1)$  is also given and thus, we can re-write the model of [16] as a special case of the following:

$$\min_{\substack{B = (B_1 \ B_2) \\ \text{rank}(B) \leq r \\ B_1 = A_1}} \|A - B\|_{GFL}. \quad (6)$$

It is obvious that, except in different norms, problem (6) is a special constrained (weighted) low-rank approximation problem as in (11).

*c) Incremental methods:* Conventional PCA [2] is an essential tool to numerically solve both RPCA and GFL problems. PCA operates at a cost of  $\min\{\mathcal{O}(m^2n), \mathcal{O}(mn^2)\}$ , which is due to the SVD of an  $m \times n$  data matrix. For RPCA algorithms, the space complexity of an SVD computation is approximately  $\mathcal{O}((m+n)r)$ , where  $r$  is the rank of the low-rank approximation matrix in each iteration, which is increasing. And a high resolution video sequence characterized by very large  $m$ , is computationally extremely expensive for the RPCA and GFL algorithms. For example, APG algorithm runs out of memory to process 600 video frames each of size  $300 \times 400$  on a computer with 3.1 GHz Intel Core i7-4770S processor and 8GB memory. In the past few decades, incremental PCA (IPCA) was developed for machine learning applications to reduce the computational complexity of performing PCA on a huge data set. The idea is to produce

an efficient SVD calculation of an augmented matrix of the form  $[A \ \tilde{A}]$  by using the SVD of  $A$ , where  $A \in \mathbb{R}^{m \times n}$  is the original matrix and  $\tilde{A}$  contains  $r$  newly added columns [17]. Similar to the IPCA, several methods have been proposed to solve the background estimation problem in an incremental manner [18], [19].

*d) Grassmannian robust adaptive subspace estimation (GRASTA):* In 2012, He et. al. [13] proposed GRASTA, a robust subspace tracking algorithm and showed its application in background modeling problems. At each time step  $i$ , GRASTA solves the following optimization problem: For a given orthonormal basis  $U_{\Omega_s} \in \mathbb{R}^{|\Omega_s| \times d}$  solve

$$\min_x \|U_{\Omega_s} x - a_i^{\Omega_s}\|_{\ell_1}, \quad (7)$$

where each video frame  $a_i \in \mathbb{R}^m$  is subsampled over the index set  $\Omega_s \subset \{1, 2, \dots, m\}$  following the model:  $a_i^{\Omega_s} = U_{\Omega_s} x + f_i^{\Omega_s} + \epsilon_{\Omega_s}$ , such that,  $x \in \mathbb{R}^d$  is a weight vector and  $\epsilon_{\Omega_s} \in \mathbb{R}^{|\Omega_s|}$  is a Gaussian noise vector. After updating  $x$ , one has to update  $U_{\Omega_s}$ .

*e) Recursive projected compressive sensing (ReProCS):*

In 2014, Guo et. al. [20] proposed another online algorithm for separating sparse and low dimensional subspace (see also [21], [22]). ReProCS is a two stage algorithm. In the first stage, given a sequence of training background frames, say  $t$ , the algorithm finds an approximate basis which is ideally of low-rank. After estimating the initial low-rank subspace in the second stage, the algorithm recursively estimates  $F_{t+1}, B_{t+1}$  and the subspace in which  $B_{t+1}$  lies.

*f) Incremental principal component pursuit (incPCP):*

Rodriguez et. al. [23] formulated the incPCP algorithm which processes one frame at a time incrementally and uses only a few frames for initialization of the prior (see also [24], [25]). incPCP follows a modified framework of conventional PCP but is built with the assumption that the partial rank  $r$  SVD of first  $k-1$  background frames  $B_{k-1}$  is known. And by using them,  $A_{k-1}$  can be written as  $A_{k-1} = B_{k-1} + F_{k-1}$ . Therefore, for a new video frame  $a_k$ , one can solve the optimization problem as follows:

$$\min_{\substack{B_k, F_k \\ \text{rank}(B_k) \leq r}} \|B_k + F_k - A_k\|_F^2 + \lambda \|F_k\|_{\ell_1}, \quad (8)$$

where  $A_k = [A_{k-1} \ a_k]$  and  $B_k = [U_r \Sigma_r V_r^T \ b_k]$  such that  $U_r \Sigma_r V_r^T$  is a partial SVD of  $B_{k-1}$ . According to [23], the initialization step can be performed incrementally.

*g) Detecting contiguous outliers in the low-rank representation (DECOLOR):* In DECOLOR, Zhou et. al. [26] combined three models: a low-rank background model, energy of the foreground support,  $S$ , generated from an Ising model, and a signal model that describes the data  $A$ , given  $B$  and  $F$ . They proposed to minimize the function:

$$\min_{\substack{B, W \\ \text{rank}(B) \leq r \\ W_{ij} \in \{0, 1\}}} \|(\mathbb{I} - W) \odot (A - B)\|_F^2 + \beta \|W\|_{\ell_1} + \gamma \|I_S, \text{vec}(W)\|_{\ell_1}, \quad (9)$$

where  $W$  is an indicator matrix whose  $(i, j)$ th entry is 1 if  $(i, j)$  represents a foreground pixel and 0 if it corresponds to

a background pixel,  $\mathbb{1}$  is the matrix of all 1s, and  $I_{S,V}$  is the incidence matrix of the graph  $\mathcal{V}$  that denotes all pixels in the foreground.

*h) Probabilistic approach to robust matrix factorization (PRMF):* In PRMF, Wang et. al. [12] decomposed  $A$  as:  $A = UV^T + F$ , where  $U$  and  $V$  are two rank  $r$  matrices. They solved the  $\ell_1$  loss function coupled with two regularizer terms:

$$\min_{U,V} \|W \odot (A - UV^T)\|_{\ell_1} + \frac{\lambda_U}{2} \|U\|_F^2 + \frac{\lambda_V}{2} \|V\|_F^2, \quad (10)$$

by using the Bayesian perspective and an expectation-maximization approach. Wang et. al. also assumed that each element of the foreground  $F$  is sampled independently from the Laplace distribution. In their model, the matrix  $W$  is an indicator matrix similar to DECOLOR and the  $\ell_2$  regularization terms prevent overfitting.

## II. WEIGHTED LOW-RANK APPROXIMATION OF MATRICES

The algorithms reviewed in the previous section tried to improve the model given by (1) via deviating from the Frobenius norm in order to encourage the detection of outliers in the data matrix. Note that the solutions to (1) as given in (2) may also suffer from another shortcoming: the fact that none of the entries of  $A$  is guaranteed to be preserved in  $X^*$ . Indeed, in many real world problems, one has good reasons to keep certain entries of  $A$  unchanged while looking for a low rank approximation. In 1987, Golub, Hoffman, and Stewart required that certain columns,  $A_1$ , of  $A$  must be preserved when one looks for a low rank approximation of  $(A_1 \ A_2)$  and considered the following *constrained* low rank approximation problem [1]: Given  $A = (A_1 \ A_2) \in \mathbb{R}^{m \times n}$  with  $A_1 \in \mathbb{R}^{m \times k}$  and  $A_2 \in \mathbb{R}^{m \times (n-k)}$ , find  $\tilde{A}_2$  such that (with  $\tilde{A}_1 = A_1$ )

$$(\tilde{A}_1 \ \tilde{A}_2) = \arg \min_{\substack{X_1, X_2 \\ \text{rank}(X_1 \ X_2) \leq r \\ X_1 = A_1}} \|(A_1 \ A_2) - (X_1 \ X_2)\|_F^2. \quad (11)$$

Inspired by applications in which  $A_1$  may contain noise, it makes more sense if we require  $\|A_1 - X_1\|_F$  small instead of looking for  $X_1 = A_1$ . This leads us to consider the following problem: Let  $\eta > 0$ , find  $(\hat{X}_1 \ \hat{X}_2)$  such that

$$(\hat{X}_1 \ \hat{X}_2) = \arg \min_{\substack{X_1, X_2 \\ \|A_1 - X_1\|_F \leq \eta \\ \text{rank}(X_1 \ X_2) \leq r}} \|(A_1 \ A_2) - (X_1 \ X_2)\|_F^2. \quad (12)$$

Or, for a large parameter  $\lambda$ , consider

$$(\hat{X}_1 \ \hat{X}_2) = \arg \min_{\substack{X_1, X_2 \\ \text{rank}(X_1 \ X_2) \leq r}} \lambda^2 \|A_1 - X_1\|_F^2 + \|A_2 - X_2\|_F^2.$$

We observed that the above problem is contained in the following more general point-wise weighted low rank approximation problem:

$$\min_{\substack{X_1, X_2 \\ \text{rank}(X_1 \ X_2) \leq r}} \|((A_1 \ A_2) - (X_1 \ X_2)) \odot (W_1 \ W_2)\|_F^2, \quad (13)$$

for  $W_1 = \lambda \mathbb{1}$  and  $W_2 = \mathbb{1}$  (a matrix whose entries are equal to 1), where  $W \in \mathbb{R}^{m \times n}$  is a weight matrix and  $\odot$  denotes the Hadamard product. This block structure in weight matrix, where very few entries are heavily weighted and most

entries stay at 1 (unweighted), is realistic in many applications. For example, in the problem of background modeling from a video sequence, each frame is a column in the data matrix and the background is a low rank (ideally of rank 1) component of the data matrix. Therefore, the weight is used to single out the columns that are more likely to be the basis of background frames and the low rank constraint enforces the search for other frames that are in the background subspace. Recent investigations in [10], [27] have shown that the above “approximately preserving” (controlled by a parameter  $\lambda$ ) weighted low rank approximation can be more effective in solving the background modeling, shadows and specularities removal, and domain adaptation problems in computer vision and machine learning. Dutta et. al. in [27] showed that if one has prior information about certain frames of the input video matrix as pure background, then one may insist on preserving those columns when looking for a low rank approximation. They reformulated the problem (13) with the weight  $W_1$  in the first block to be chosen entry-wise rather than scaled by a single constant  $\lambda$ . This gives more freedom to pixel-wise control of the columns of the given matrix to be preserved. In this paper, (i) we provide a consolidated treatment of our recent contributions to the background modeling problem previously proposed in several conference proceeding papers based on the method that can be categorized as weighted low rank approximation of matrices; and (ii) we present more systematical experiments and a thorough comparison of our proposed algorithms against several state-of-the-art background modeling methods on the mainstream data-sets including Stuttgart, I2R, Wallflower, CDNet 2014, and the SBI data-set, [28], [29], [30], [31], [32], [33].

Here is the more detailed outline of the rest of the paper.

We first discuss an algorithm (Algorithm 1) to solve (13) numerically when  $(W_1)_{ij} \in [\alpha, \beta]$  and  $W_2 = \mathbb{1}$ . This serves two purposes: that i) our algorithm takes advantage of the block structure of the weights that results in efficient numerical performance as compared to the existing algorithms for solving the general weighted low-rank approximation problem [34], and that ii) we have detailed convergence analysis for the algorithm.

Next, we review a batch background estimation model in Algorithm 2 that is built on Algorithm 1. More specifically, unlike the models described previously (such as in GFL and ReProCS methods reviewed in section I-B), we do not require any training frames but our algorithm can learn the background frame indices robustly.

Finally, we investigate an adaptive batch-incremental model (Algorithm 3). Our model finds the background frame indices robustly and incrementally to process the entire video sequence adaptively by going through a sequence of small batches. Therefore, unlike Algorithm 2, Algorithm 3 does not use the entire sequence at once. Rather it operates on local small batch of video frames and use the background information in nearby frames. As a result, it is more time and memory efficient.

We note that similar to the  $\ell_1$  norm used in conventional and in the incremental methods, a weighted Frobenius norm used in Algorithm 1 makes Algorithm 2 and 3 robust to the

outliers for the background modeling problems. Our empirical validation shows that a properly weighted Frobenius norm can perform as well as or better than that of state-of-the-art  $\ell_1$  norm minimization (or any other norms) algorithms for background estimation.

To conclude the paper, we present extensive numerical experiments to show that our batch-incremental model is as fast as incPCP and ReProCS, also, it can deal with high quality video sequences as accurately as incPCP and ReProCS. Some conventional algorithms, such as supervised GFL or ReProCS require an initial training sequence which does not contain any foreground object. Our experimental results for both synthetic and real video sequences show that unlike the supervised GFL and ReProCS, our models does not require a prior. We believe that the adaptive nature of our algorithm is well suited for real time high-definition video surveillance.

### III. WEIGHTED LOW-RANK APPROXIMATION: THEORY

**A**S in the standard low rank approximation, the constrained low rank approximation problem (11) of Golub, Hoffman, and Stewart has a closed form solution which is given in the following theorem:

*Theorem 1:* [1] Assume  $\text{rank}(A_1) = k$  and  $r \geq k$ , the solution  $\tilde{A}_2$  in (11) is given by

$$\tilde{A}_2 = P_{A_1}(A_2) + H_{r-k}(P_{A_1}^\perp(A_2)), \quad (14)$$

where  $P_{A_1}$  and  $P_{A_1}^\perp$  are the projection operators to the column space of  $A_1$  and its orthogonal complement, respectively.

Let  $(\tilde{X}_1(W) \ \tilde{X}_2(W))$  be a solution to (13). Let  $\lambda_j = \min_{1 \leq i \leq m} (W_1)_{ij}$  and  $\lambda = \min_{1 \leq j \leq k} \lambda_j$ . Denote  $\mathcal{A} = P_{A_1}^\perp(A_2)$  and  $\tilde{\mathcal{A}} = P_{\tilde{X}_1(W)}^\perp(A_2)$ . Also denote  $s = \text{rank}(\mathcal{A})$  and let the ordered non-zero singular values of  $\mathcal{A}$  be  $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_s > 0$ .

*Theorem 2:* Suppose that  $\sigma_{r-k} > \sigma_{r-k+1}$ . Then we have, as  $\lambda \rightarrow \infty$  and  $W_2 = \mathbb{I}$ ,

$$(\tilde{X}_1(W) \ \tilde{X}_2(W)) = A_G + O\left(\frac{1}{\lambda}\right),$$

where  $A_G = (A_1 \ \tilde{A}_2)$  is defined to be the unique solution to (11).

*Remark 3:* (i) We note that  $A_G$  is unique due to the assumption  $\sigma_{r-k} > \sigma_{r-k+1}$  [1]. (ii) The convergence  $(\tilde{X}_1(W) \ \tilde{X}_2(W)) \rightarrow A_G$  alone is expected and indeed implied by a general result in [35]. However, it does not specify a rate of convergence.

#### A. A brief literature review

As it turns out, (13) can be viewed as a generalized total least squares problem (GTLS) and can be solved in closed form as a special case of weighted low rank approximation with a rank-one weight matrix by using a SVD of the given matrix  $(\lambda A_1 \ A_2)$  [36], [37]. As a consequence of the closed form solutions, one can verify that the solution to (11) is the limit case of the solutions to (13) as  $\lambda \rightarrow \infty$ . Thus, (11) can be viewed as a special case when “ $\lambda = \infty$ ”. A careful reader may also note that, problem (13) can be cast as a special case of structured low rank problems with element-wise weights [35], [38].

The weighted low rank approximation problem was studied first when  $W$  is an indicator weight for dealing with the missing data case ([39], [40]) and then for more general weight in machine learning, collaborative filtering, 2-D filter design, and computer vision [41], [42], [43], [44], [45], [46]. For example, if SVD is used in quadrantly-symmetric two-dimensional (2-D) filter design, as explained in [44] (see also [45], [46]), it might lead to a degraded construction in some cases as it is not able to discriminate between the important and unimportant components of  $X$ . To address this problem, a weighted least squares matrix decomposition method was first proposed by Shpak [46]. Following his idea of assigning different weights to discriminate between important and unimportant components of the test matrix, Lu, Pei, and Wang ([45]) designed a numerical procedure to solve (13) for general weight  $(W_1 \ W_2)$ .

*Remark 4:* There is another formulation of weighted low rank approximation problem defined as in [44]:

$$\min_{X \in \mathbb{R}^{m \times n}} \|A - X\|_Q^2, \quad \text{subject to } \text{rank}(X) \leq r, \quad (15)$$

where  $Q \in \mathbb{R}^{mn \times mn}$  is a symmetric positive definite weight matrix. Denote  $\|A - X\|_Q^2 := \text{vec}(A - X)^T Q \text{vec}(A - X)$ , where  $\text{vec}(\cdot)$  is an operator which maps the entries of  $\mathbb{R}^{m \times n}$  to  $\mathbb{R}^{mn \times 1}$ . It is easy to see that (13) is a special case of (15) with a diagonal  $Q$ . In this paper, we will not use this more general formulation for simplicity.

#### B. Why the problem becomes more difficult to solve when weights present

Note that, in general, the weighted low rank approximation problem (13) does not have a closed form solution, even when  $W_2 = \mathbb{I}$  [41], [44]. With  $W_2 = \mathbb{I}$ , we can write (13) as

$$\min_{\substack{X_1, X_2 \\ \text{rank}(X_1 \ X_2) \leq r}} (\|(A_1 - X_1) \odot W_1\|_F^2 + \|A_2 - X_2\|_F^2).$$

Let  $\text{rank}(X_1) = k$ . It is easy to see that any  $X_2$  such that  $\text{rank}(X_1 \ X_2) \leq r$  can be given in the form

$$X_2 = X_1 C + B D,$$

for some arbitrary matrices  $B \in \mathbb{R}^{m \times (r-k)}$ ,  $D \in \mathbb{R}^{(r-k) \times (n-k)}$ , and  $C \in \mathbb{R}^{k \times (n-k)}$ . Hence we need to solve

$$\min_{X_1, C, B, D} (\|(A_1 - X_1) \odot W_1\|_F^2 + \|A_2 - X_1 C - B D\|_F^2). \quad (16)$$

Note that, using a block structure, we can write (16) as (with a special low rank structure):

$$\min_{X_1, C, B, D} \left\{ \left\| \begin{pmatrix} (A_1 \ A_2) - (X_1 \ B) \begin{pmatrix} I_k & C \\ 0 & D \end{pmatrix} \end{pmatrix} \odot (W_1 \ \mathbb{I}) \right\|_F^2 \right\},$$

which is in a form of the alternating weighted least squares problem in the literature [41], [36]. But we will not follow the general algorithm proposed in [36] for the following reasons: that (i) due to the special structure of the weight, our algorithm is more efficient than [36] (see Algorithm 3.1, in p. 42 [36]), that (ii) it allows a detailed convergence analysis which is usually not available in other algorithms proposed in the literature [41], [44], [36], and that (iii) it can handle bigger

size matrices as we will demonstrate in the numerical result section. If  $k = 0$ , then (16) reduces to an unweighted rank  $r$  factorization of  $A_2$  and can be solved as an alternating least squares problem [47], [48].

### C. Solving the weighted problem

Denote  $F(X_1, C, B, D) = \|(A_1 - X_1) \odot W_1\|_F^2 + \|A_2 - X_1 C - B D\|_F^2$  as the objective function. The function  $F$  is minimized by using an alternating strategy [9], [49] of minimizing the function with respect to each component iteratively:

$$\begin{cases} (X_1)_{p+1} = \arg \min_{X_1} F(X_1, C_p, B_p, D_p), \\ C_{p+1} = \arg \min_C F((X_1)_{p+1}, C, B_p, D_p), \\ B_{p+1} = \arg \min_B F((X_1)_{p+1}, C_{p+1}, B, D_p), \\ \text{and, } D_{p+1} = \arg \min_D F((X_1)_{p+1}, C_{p+1}, B_{p+1}, D). \end{cases} \quad (17)$$

Note that each of the minimizing problem for  $X_1, C, B$ , and  $D$  can be solved explicitly by looking at the gradients of  $F(X_1, C, B, D)$ . But finding an update rule for  $X_1$  turns out to be more involved than the other three variables due to the interference of the weight  $W_1$ . We update  $X_1$  element wise along each row. Therefore we will use the notation  $X_1(i, :)$  to denote the  $i$ -th row of the matrix  $X_1$ . The numerical process is described in Algorithm 1.

---

#### Algorithm 1: WLR Algorithm

---

```

1 Input :  $A = (A_1 \ A_2) \in \mathbb{R}^{m \times n}$  (the given matrix);
            $W = (W_1 \ \mathbb{1}) \in \mathbb{R}^{m \times n}$  (the weight), threshold
            $\epsilon > 0$ ;
2 Initialize:  $(X_1)_0, C_0, B_0, D_0$ ;
3 while not converged do
4    $E_p = A_1 \odot W_1 \odot W_1 + (A_2 - B_p D_p) C_p^T$ ;
5   for  $i = 1 : m$  do
6      $(X_1(i, :))_{p+1} = (E(i, :))_p (\text{diag}(W_1^2(i, 1))$ 
        $W_1^2(i, 2) \cdots W_1^2(i, k)) + C_p C_p^T)^{-1}$ ;
   end
7    $C_{p+1} = ((X_1)_{p+1}^T (X_1)_{p+1})^{-1} (X_1)_{p+1}^T (A_2 - B_p D_p)$ ;
8    $B_{p+1} = (A_2 - (X_1)_{p+1} C_{p+1}) D_p^T (D_p D_p^T)^{-1}$ ;
9    $D_{p+1} = (B_{p+1}^T B_{p+1})^{-1} B_{p+1}^T (A_2 - (X_1)_{p+1} C_{p+1})$ ;
10   $p = p + 1$ ;
11 end
12 Output :  $(X_1)_{p+1}, (X_1)_{p+1} C_{p+1} + B_{p+1} D_{p+1}$ .
```

---

### D. Convergence

In this section we comment on the convergence on Algorithm 1. We quote the convergence results of Algorithm 1 without their proof. For detailed proofs we refer to [34]. The first theorem shows a relation that involves the iterates and the reconstruction error of the objective function.

**Theorem 5:** For a fixed  $(W_1)_{ij} > 0$  let  $m_p = F((X_1)_p, C_p, B_p, D_p)$  for  $p = 1, 2, \dots$ . Then,

$$\begin{aligned} m_p - m_{p+1} = & \|((X_1)_p - (X_1)_{p+1}) \odot W_1\|_F^2 \\ & + \|((X_1)_p - (X_1)_{p+1}) C_p\|_F^2 \\ & + \|(X_1)_{p+1} (C_p - C_{p+1})\|_F^2 \\ & + \|(B_p - B_{p+1}) D_p\|_F^2 \\ & + \|B_{p+1} (D_p - D_{p+1})\|_F^2. \end{aligned} \quad (18)$$

**Remark 6:** From Theorem 5 we know that the non-negative sequence  $\{m_p\}$  is non-increasing. Therefore,  $\{m_p\}$  has a limit.

By using Theorem 5 we have the following estimates.

**Corollary 7:** We have

- (i)  $m_p - m_{p+1} \geq \frac{1}{2} \|B_{p+1} D_{p+1} - B_p D_p\|_F^2$  for all  $p$ .
- (ii)  $m_p - m_{p+1} \geq \|((X_1)_p - (X_1)_{p+1}) \odot W_1\|_F^2$  for all  $p$ .

Consider the situation when

$$\sum_{p=1}^{\infty} \sqrt{m_p - m_{p+1}} < +\infty. \quad (19)$$

**Theorem 8:**

- (i) We have the following:  $\sum_{p=1}^{\infty} \|B_{p+1} D_{p+1} - B_p D_p\|_F^2 < \infty$ , and  $\sum_{p=1}^{\infty} (\|((X_1)_p - (X_1)_{p+1}) \odot W_1\|_F^2) < \infty$ .
- (ii) If (19) holds then  $\lim_{p \rightarrow \infty} B_p D_p$  and  $\lim_{p \rightarrow \infty} (X_1)_p$  exist. Furthermore if we write  $L^* := \lim_{p \rightarrow \infty} B_p D_p$  then  $\lim_{p \rightarrow \infty} B_{p+1} D_p = L^*$  for all  $p$ .

Note that Corollary 8 only states the convergence of the sequence  $\{B_p D_p\}$  but not of  $\{B_p\}$  and  $\{D_p\}$  separately. The convergence of  $\{B_p\}$  and  $\{D_p\}$  can be obtained with stronger assumption as demonstrated in the next result.

**Theorem 9:** Assume (19) holds.

- (i) If  $B_p$  is of full rank and  $B_p^T B_p \geq \gamma I_{r-k}$  for large  $p$  and some  $\gamma > 0$  then  $\lim_{p \rightarrow \infty} D_p$  exists.
- (ii) If  $D_p$  is of full rank and  $D_p D_p^T \geq \delta I_{r-k}$  for large  $p$  and some  $\delta > 0$  then  $\lim_{p \rightarrow \infty} B_p$  exists.
- (iii) If  $X_1^* := \lim_{p \rightarrow \infty} (X_1)_p$  is of full rank, then  $C^* := \lim_{p \rightarrow \infty} C_p$  exists. Furthermore, if we write  $L^* = B^* D^*$ , for  $B^* \in \mathbb{R}^{m \times (r-k)}$ ,  $D^* \in \mathbb{R}^{(r-k) \times (n-k)}$ , then  $(X_1^*, C^*, B^*, D^*)$  will be a stationary point of  $F$ .

## IV. WEIGHTED LOW-RANK APPROXIMATION: APPLICATION TO BACKGROUND MODELING

**W**E focus on two algorithms for background modeling by using Algorithm 1. Both algorithms robustly determine the frame indices which are contenders of the background frames (with least or no foreground movement) and then take the advantage of weight matrix  $W$  to estimate the background and the foreground.

---

#### Algorithm 2: Background Estimation using WLR

---

```

1 Input :  $A = (A_1 \ A_2) \in \mathbb{R}^{m \times n}$  (the given matrix);
            $W = (W_1 \ \mathbb{1}) \in \mathbb{R}^{m \times n}$  (the weight), threshold
            $\epsilon > 0$ ,  $i_1, i_2 \in \mathbb{N}$ ;
2 Run PCA to get low rank  $B_{In}$  and  $F_{In} = A - B_{In}$ ;
3 Learn background frame indices  $S$  from  $B_{In}$  and  $F_{In}$ ;
4 Set  $k = \lceil |S|/i_1 \rceil$ ,  $r = k + i_2$ ;
5 Rearrange data:  $\tilde{A}_1 = (A(:, i))_{m \times k}$ , randomly chosen  $k$  frames
   from  $i \in S$  and  $\tilde{A}_2 = (A(:, i'))_{m \times (n-k)}$ ,  $i'$  from the
   remaining frames;
6 Apply WLR on  $\tilde{A} = (\tilde{A}_1 \ \tilde{A}_2)$  to obtain  $\tilde{X}$ ;
7 Rearrange the columns of  $\tilde{X}$  similar to  $A$  to find  $X$ ;
8 Output :  $X$ .
```

---

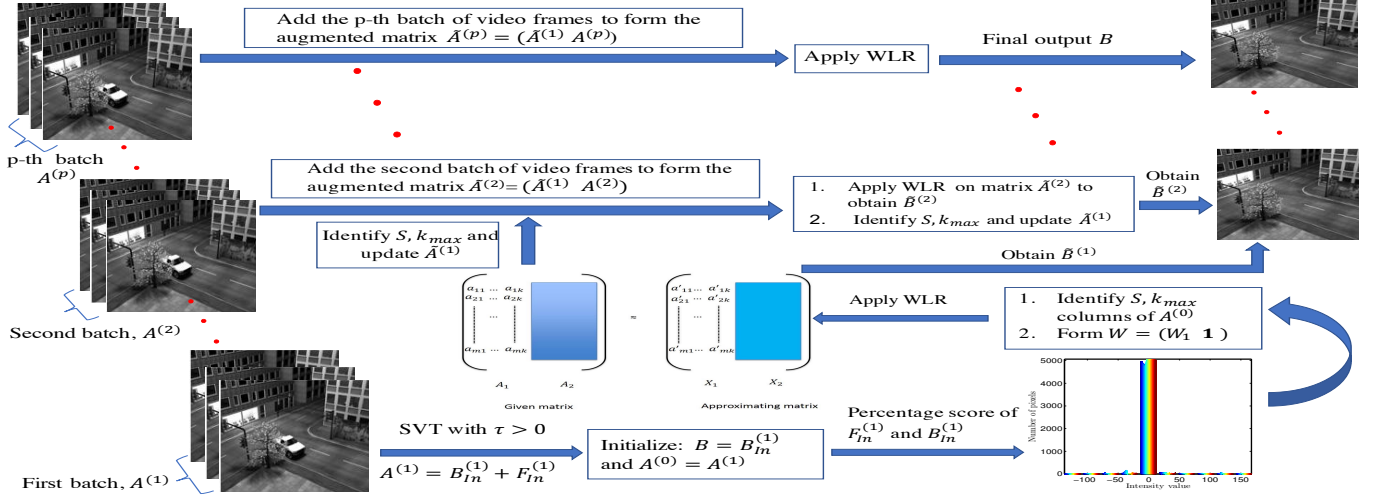


Fig. 1: A flowchart for WLR inspired batch-incremental background estimation model proposed in Algorithm 3.

### Algorithm 3: Incremental Background Estimation using WLR (inWLR)

- 1 **Input** :  $p, A = (A^{(1)} A^{(2)} \dots A^{(p)}) \in \mathbb{R}^{m \times n}$ ,  $\tau > 0$  (for SVT),  $\alpha, \beta > 0$  (for weights), threshold  $\epsilon > 0$ ,  $k_{\max}, i_r \in \mathbb{N}$ ;
- 2 Run SVT on  $A^{(1)}$  with parameter  $\tau$  to obtain:  
 $A^{(1)} = B_{In}^{(1)} + F_{In}^{(1)}$ ;
- 3 Initialize the background block by  $B = B_{In}^{(1)}$  and  $A^{(0)} = A^{(1)}$ ;
- 4 **for**  $j = 1 : p$  **do**
- 5   Identify the indices  $S$  of at most  $k_{\max}$  columns of  $A^{(j-1)}$  that are closest to background using  $B$  and  $F = A^{(j-1)} - B$ ;
- 6   Set  $k = \#(S)$ ,  $r = k + i_r$ ;
- 7   Set the first block:  $\tilde{A}_1 = (A^{(j-1)}(:, i))_{m \times k}$  with  $i \in S$ ;
- 8   Define  $W = (W_1 \mathbb{1})$  with  $W_1 \in \mathbb{R}^{m \times k}$  where  $(W_1)_{ij}$  are randomly chosen from  $[\alpha, \beta]$ ;
- 9   Apply **Algorithm 1** on  $\tilde{A}^{(j)} = (\tilde{A}_1 A^{(j)})$  using threshold  $\epsilon$  and weight  $W$  to obtain its low rank component  $\tilde{B}^{(j)}$  and define  $\tilde{F}^{(j)} = \tilde{A}^{(j)} - \tilde{B}^{(j)}$ ;
- 10   Take the sub-matrix of  $\tilde{B}^{(j)}$  corresponding to the  $A^{(j)}$  block such that  $A^{(j)} = B^{(j)} + F^{(j)}$ ;
- 11   Update the background block:  $B = \tilde{B}^{(j)}$ ;
- end**
- 12 **Output** :  $B = (B^{(1)}, B^{(2)}, \dots, B^{(p)})$ .

#### A. Background estimation by using Algorithm 2

First, we describe a batch background modeling algorithm. To initialize, we first solve WLR (by using Algorithm 1) for  $W = I_n$  (this is just PCA) to obtain a coarse estimate of the background and foreground:  $A = B_{In} + F_{In}$ , where  $B_{In}$  is a low rank approximation to  $A$  given by PCA. Next, we use  $B_{In}$  and  $F_{In}$  to learn the frame indices that are closest to the pure background (with least or no foreground movements). This is done heuristically (as in [10](see Figure 2)). By setting a threshold  $\epsilon_1 > 0$  based on the histogram of  $F_{In}$ , we convert  $F_{In}$  into a binary matrix  $LF_{In}$ : all entries of  $F_{In}$  that are bigger than  $\epsilon_1$  are replaced by 1 and the others are replaced by 0. The matrix  $B_{In}$  is directly converted to a binary matrix  $LB_{In}$ . Next, we calculate the ratios of the frame sum (i.e.

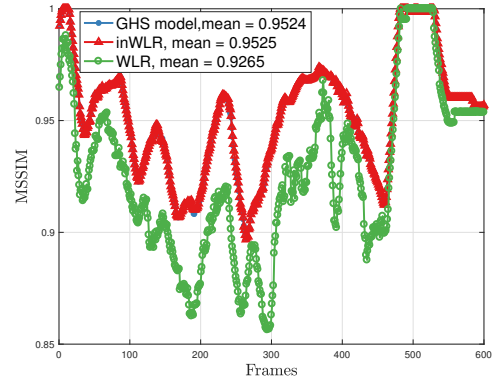


Fig. 2: Comparison of MSSIM of WLR acting on all frames, inWLR ( $p = 6$ ), and GHS inspired background estimation model with resolution [144, 176].

the column sum) of  $LF_{In}$  to the corresponding frame sum of  $LB_{In}$  and identify the indices with ratios less than the mode of these ratios as possible pure background frame indices. Finally, we apply WLR by putting the weight at the learned frame indices to decompose the data matrix  $A$  into background and foreground:  $A = B + F$ . Our experiments show that the performance depends more on the correct location (indices) of the background frames than on the values of the weight. We remark that Dutta and Li [27] and Xin et al. [16] used the pure background frames in their background estimation model, assuming the frames were already given. On the contrary, Algorithm 2 learns the background frame indices from the data, thus providing a robust background estimation model.

#### B. An incremental model using WLR

Next we propose an incremental weighted low rank approximation (inWLR) algorithm for background modeling (see Algorithm 3 and Figure 1). This algorithm proves to be efficient compared to Algorithm 2 when the video sequence contains a lot of frames. Our algorithm fully exploits WLR, in which a



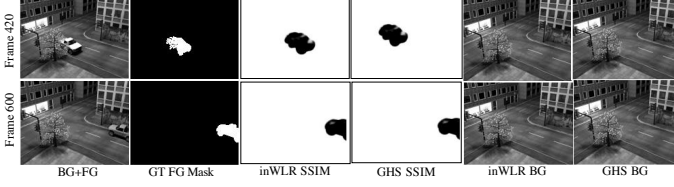


Fig. 3: SSIM map of inWLR and GHS inspired background estimation model. Top to bottom: Frame 420 with dynamic foreground, frame 600 with static foreground. SSIM index of the methods are 0.95027 and 0.96152, respectively.

prior knowledge of the background space can be used as an additional constraint to obtain the low rank (thus the background) estimation of the data matrix  $A$ . First, we partition the original video sequence into  $p$  batches:  $A = (A^{(1)} \ A^{(2)} \ \dots \ A^{(p)})$ , where the batch sizes do not need to be equal. Instead of working on the entire video sequence, the algorithm incrementally works through each batch. To initialize, the algorithm coarsely estimates the possible background frame indices of  $A^{(1)}$ : we run the classic singular value thresholding (SVT) of Cai et. al. [50] on  $A^{(1)}$  (we can afford this since the size of  $A^{(1)}$  is much smaller than that of  $A$ ) to obtain a low rank component (containing the estimations of background frames)  $B_{In}^{(1)}$  and let  $F_{In}^{(1)} = A^{(1)} - B_{In}^{(1)}$  be the estimation of the foreground matrix (Step 2). From the above estimates, we obtain the initialization for  $B$  and  $A^{(0)}$  (Step 3). Then, we go through each batch  $A^{(j)}$ , using the estimates of the background from the previous batch as prior for the WLR algorithm to obtain the background  $\hat{B}^{(j)}$  (Step 9). To determine the indices of the frames that contain the least information of the foreground we identify the “best background frames” by using a modified version of the percentage score model by Dutta et. al. [10] (Step 5). Using this modified model allows us to estimate  $k$ ,  $r$ , and the first block  $\hat{A}_1$  which contains the background prior knowledge (Steps 6-7). Weight matrix  $W = (W_1 \ \mathbb{1})$  is chosen by randomly picking the entries of the first block  $W_1$  from an interval  $[\alpha, \beta]$  by using an uniform distribution, where  $\beta > \alpha > 0$  are large (Step 8). To understand the effect of using a large weight in  $W_1$ , we refer the reader to Theorem 2 and [27], [34]. Finally, we collect background information for next iteration (Steps 10-11). Note that the number of columns of the weight matrix  $W_1$  is  $k$ , which is controlled by bound  $k_{max}$  so that the column size of  $\hat{A}^{(j)}$  does not grow with  $j$ . The output of the algorithm is the background estimations for each batch collected in a single matrix  $B$ . When the camera motion is small, updating the first block matrix  $\hat{A}_1$  (Step 7) has trivial impact on the model since it does not change much. However, when the background is continuously evolving (but slowly), our inWLR could be proven very robust as new frames are entering in the video. Moreover, we show that Algorithm 3 is faster compare with Algorithm 2.

### C. Complexity Analysis

Now, we analyze the complexity of Algorithm 3 for equal batch size and compare it with Algorithm 2. Primarily, the cost of the SVT algorithm in Step 2 is only  $\mathcal{O}(\frac{mn^2}{p^2})$ . Next,

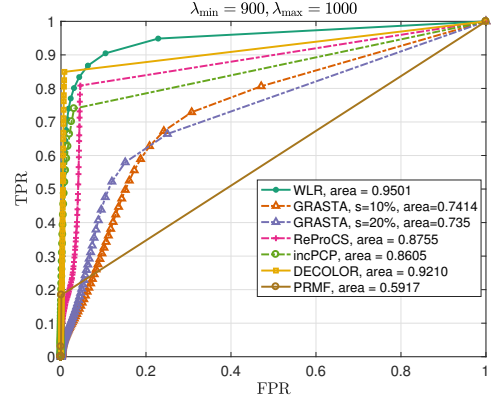


Fig. 4: ROC curve to compare between WLR, GRASTA, ReProCS, incPCP, DECOLOR, and PRMF on Basic video, resolution  $144 \times 176$ .

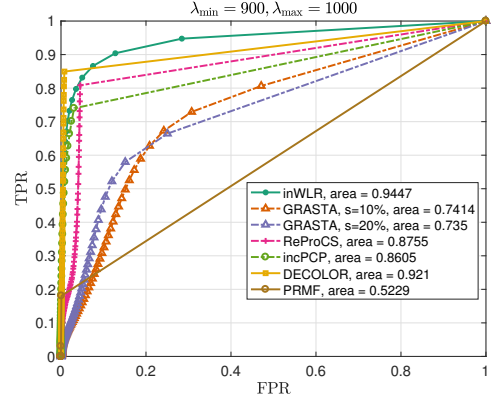


Fig. 5: ROC curve to compare between inWLR, GRASTA, ReProCS, incPCP, DECOLOR, and PRMF on Basic video, resolution  $144 \times 176$ .

in Step 9, the complexity of implementing Algorithm 1 is  $\mathcal{O}(mk^3 + \frac{mnr}{p})$ . Note that  $r$  and  $k$  are linearly related and  $k \leq k_{max}$ . Once we obtain a refined estimate of the background frame indices  $S$  as in Step 5 and form an augmented matrix by adding the next batch of video frames, a very natural question in proposing our WLR inspired Algorithm 3 is: why do we use Algorithm 1 in each incremental step (Step 9) of Algorithm 3 instead of using a closed form solution (14) of GHS? We justify as follows: the estimated background frames  $\hat{A}_1$  are not necessarily exact background; they are only estimations of background. Thus, GHS inspired model may be forced to follow the wrong data while inWLR allows enough flexibility to find the best fit to the background subspace. This is confirmed by our numerical experiments (see Section V and Figure 2). Thus, to analyze the entire sequence in  $p$  batches, the complexity of Algorithm 3 is approximately  $\mathcal{O}(m(k^3p + nr))$ . Note that the complexity of Algorithm 3 is dependent on the partition  $p$  of the original data matrix. Our numerical experiments suggest that for video frames of varying sizes, the choice of  $p$  plays an important role and is empirically determined. Unlike Algorithm 3, if Algorithm 2 is used on the entire data set and if the number of possible background frame



Dataset	Video	Frames	Resolution used
Stuttgart [28]	Basic	600	144 × 176
Wallflower [30]	Waving Tree	66	120 × 160
SBI [32]	Snellen	321	144 × 176
	IBMTes2	90	144 × 144
	HumanBody	740	144 × 144
	Foliage	300	144 × 176
	Candela	350	144 × 176
	Caviar2	460	144 × 176
CDnet 2014 [31]	Hall and Monitor	296	144 × 176
	Abandoned Box	300	144 × 176
	Backdoor	300	144 × 176
	Busstation	600	144 × 176
	Intermittent Pan	400	144 × 176
	Tunnel Exit	300	144 × 176
	Port	1000	144 × 176
	Fountain	300	144 × 176
	Overpass	1100	144 × 176
	Fountain 2	1100	144 × 176
	Canoe	1100	144 × 176
	Fall	1100	144 × 176
I2R/Li dataset [29]	Meeting Room	1209	64 × 80
	Watersurface	162	128 × 160
	Campus	600	64 × 80
	Fountain	500	64 × 80

TABLE I: Data used in this paper.

indices is  $k'$ , then the complexity is  $\mathcal{O}(mk'^3 + mnk')$ . When  $k'$  grows with  $n$  and becomes much bigger than  $k_{max}$  in order to achieve competitive performance, we see that Algorithm 1 tends to slow down with higher overhead than Algorithm 3 does.

## V. EXPERIMENTS

**W**E present our extensive numerical experiments and show the effectiveness of our background modelling algorithms on synthetic and real world video sequences and compare them with several state-of-the-art background modelling algorithms.

### A. Data sets

We extensively use 24 gray scale videos from the Stuttgart, I2R, Wallflower, CDNet 2014, and the SBI dataset [28], [29], [32], [33], [30], [31]. Stuttgart is a synthetic dataset, but all other datasets are real world videos. They contain several challenges, for example, static foreground, dynamic background, change in illumination, occlusion and disocclusion of static and dynamic foreground. We refer the readers to Table I to get an overall idea of the number of frames of each video sequence used, video type, and resolution.

### B. Metrics used for quantitative comparison

We use four different metrics for this purpose: traditionally used receiver and operating characteristic (ROC) curve, peak signal to noise ratio (PSNR), and the most advanced measures mean structural similarity index (MSSIM), and multiscale structural similarity index (MSSSIM)[53], [54], [55] as they mostly agree with the human visual perception [53]. When a ground truth mask (foreground or background) is available for each video frame, we use a pixel-based measure of  $F$  or  $B$ , the foreground or background recovered by each method

to form the confusion matrix for the ROC predictive analysis. In our case, the pixels are represented by 8 bits per sample, and  $M_I$ , the maximum pixel value of the image is 255. Therefore, a uniform threshold vector  $\text{linspace}(0, M_I, 100)$  is used to compare the pixelwise predictive analysis between each recovered foreground or background frame and the corresponding ground truth. PSNR is calculated by using the metric  $10\log_{10}\frac{M_I^2}{MSE}$ , such that  $MSE = \frac{1}{mn}\|G(:,i) - R(:,i)\|_2^2$ , where  $R(:,i)$  is the recovered vectorized BG/FG frame and  $G(:,i)$  is the corresponding vectorized GT frame. To calculate the SSIM and MSSSIM of each recovered FG/BG video frame, we consider a  $11 \times 11$  Gaussian window with standard deviation ( $\sigma$ ) 1.5 unless otherwise specified. We consider the area covered by the ROC curve of an algorithm in a unit square as a measure of its performance, where the higher the value is the better. Similarly, for SSIM and MSSSIM the values that are closer to 1 are better. Lastly, the PSNR of a reconstructed image generally falls in the range 30-50 dB, where the higher the better as well.

### C. Results on Basic scene of Stuttgart artificial video

Due to the availability of ground truth frames for each foreground mask, we use 600 frames of the Basic scenario of the Stuttgart artificial video sequence [28] for quantitative and qualitative analysis. To capture an unified comparison against each method, we resize the video frames to  $144 \times 176$  and for inWLR set  $p = 6$ ; that is, we add a batch of 100 new video frames in every iteration until all frames are exhausted.

1) *Comparison with GHS*: Because the Basic scenario has no noise, once we estimate the background frames, GHS can be used as a baseline method for comparing the effectiveness of Algorithm 3. To demonstrate the benefit of using an iterative process as inWLR (Algorithm 1), we first compare the performance of Algorithm 3 against the GHS inspired model. We also compare Algorithm 2 acting on all 600 frames with the parameters specified in [51]. We use MSSIM to quantitatively evaluate the overall image quality. To calculate MSSIM we perceive the information of how the high-intensity regions of the images come through the noise, and consequently, we pay much less attention to the low-intensity regions. We remove the noisy components  $F$  by thresholding it by  $\epsilon_1$ . Figure 2 indicates that the inWLR and GHS inspired model produce the same result, but inWLR is more time efficient than GHS. To process 600 frames, inWLR takes **18.06** seconds. In contrast, GHS inspired model takes 160.17 seconds and WLR takes **39.99** seconds. In Figure 3, the SSIM map of two sample video frames of the Basic scenario show that both methods recover the similar quality background.

2) *Comparison with RPCA*: Now we compare Algorithm 2 and Algorithm 3 with the RPCA. For this purpose we consider APG and iEALM. For both algorithms we set  $\lambda = 1/\max\{m, n\}$ , and for iEALM we choose  $\mu = 1.25/\|A\|_2$  and  $\rho = 1.5$ , where  $\|A\|_2$  is the spectral norm (maximum singular value) of  $A$ . From Figure 6 it is clear that when the foreground is static for a few frames both APG and iEALM can not remove the static foreground object. Moreover, to process 600 frames of resolution  $144 \times 176$ , iEALM and APG took 501.46 seconds and 572.49 seconds, respectively.

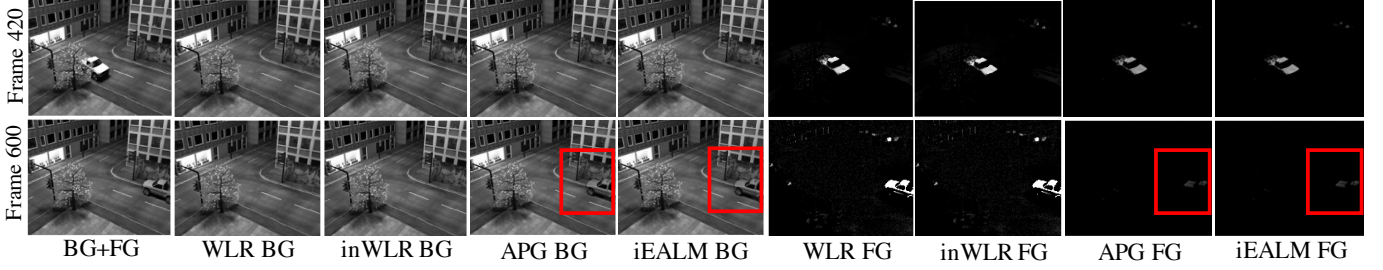


Fig. 6: Qualitative comparison of WLR and inWLR with iEALM and APG. Top to bottom: Frame 420 with dynamic foreground, frame 600 with static foreground.

Algorithm	Appearing in	Reference
Weighted Low-rank approximation (WLR)	Figure 2, 4, 6, 7, 9, and 13	This paper (Algorithm 2), [51]
Incremental Weighted Low-rank approximation (inWLR)	Figure 1, 2, Figure 5-12, Figure 14-16, Table III, and IV	This paper (Algorithm 3), [52]
Inexact Augmented Lagrange Method of Multipliers (iEALM)	Figure 6	[9]
Accelerated Proximal Gradient (APG)	Figure 6	[8]
Goulb et. al. inspired BG model	Figure 2, 3	[1], [51]
Supervised Generalized Fused Lasso (GFL)	Figure 10, 14	[16]
Grassmannian Robust Adaptive Subspace Tracking (GRASTA)	Figure 4, 5, 9	[13]
Recursive Projected Compressive Sensing (ReProCS)	Figure 4, 5, 7, 8, 14	[20], [21], [22]
Incremental Principal Component Pursuit (incPCP)	Figure 4, 5, 7, 9, 10	[23], [24], [25]
Probabilistic Robust Matrix Factorization (PRMF)	Figure 4, 5, and 10	[12]
Detecting Contiguous Outliers in the Low-Rank Representation (DECOLOR)	Figure 4, 5, and 10	[26]

TABLE II: Algorithms compared in this paper.

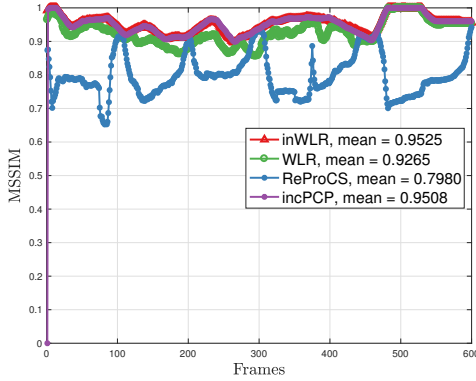


Fig. 7: Mean SSIM to compare between WLR, inWLR, ReProCS, and incPCP on Basic video, frame size  $144 \times 176$ .

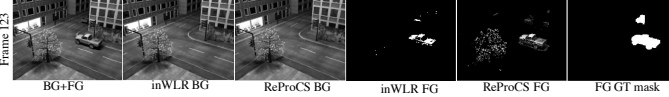


Fig. 8: Basic scenario frame 123. Both methods recover similar quality background, however, ReProCS foreground has more false positives than inWLR. This explains why ReProCS suffers in quantitative evaluation.

Video	MSSIM	MSSSIM	PSNR	Area under ROC curve
Snellen	0.99179	0.99891	60.9451	0.9525
IBMtest2	0.99979	0.99998	76.2776	0.9985
Human Body	0.9994	0.9999	71.5218	-
Foliage	0.9865	0.99803	58.5055	-
Candela	0.9999	0.999995	80.2983	0.9988
Caviar2	0.99998	0.99999	87.8575	-
HallMonitor	0.99993	0.99998	84.8487	1

TABLE III: Performance of inWLR on SBI dataset [32].

3) *Comparison with other state-of-the-art-methods:* Next we compare Algorithm 2 and Algorithm 3 with other state-of-the-art background estimation methods, such as, GRASTA (with different subsample ratio), incPCP, ReProCS, PRMF, GFL, and DECOLOR. We do not include GOSUS [14] as it is very similar to GRASTA. For GRASTA, we set the subsample percentage  $s$  at 0%, 10%, 20%, and at 30% respectively, estimated rank 60, and keep the other parameters the same as those in [13]. We use 200 background frames of the Basic sequence for initialization of ReProCS. According to [23], the initialization step can be performed incrementally. For the Stuttgart sequence, the algorithm uses the first video frame for initialization. PRMF and DECOLOR are unsupervised algorithms. We set the target rank for PRMF to 5 and the other parameters are kept same as in the software package. For DECOLOR we use the static camera interface of the code and the parameters are kept same as they are mentioned in the software package. The ROC curves in Figures 4 and 5 to demonstrate that our proposed algorithms outperform other methods. In Figure 7 we present the mean SSIM of the recovered foreground frames and Algorithm 3 outperforms ReProCS. In contrast, incPCP has similar mean SSIM. Moreover, in Figure 9 all methods except GRASTA appear to perform equally well on the Basic scenario. However, when the foreground is static (as in frames 551-600 of the Basic scenario) neither incPCP, PRMF, nor, DECOLOR can capture the static foreground object, thus result the presence of the static car as a part of the background (see Figure 10). For supervised GFL model, we use 200 frames from each scenario (Basic and Waving Tree) for training purpose. The background recovered and the SSIM map in Figures 10 show that GFL provides a comparable reconstruction and can effectively remove the static car (also see Figure 14). However,

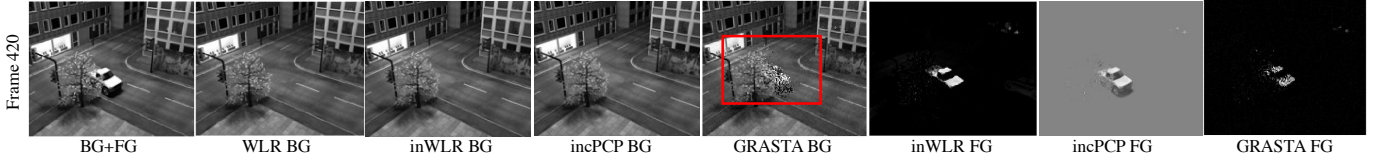


Fig. 9: Basic scenario frame 420. GRASTA with subsample rate 10% recovers a fragmentary foreground and degraded background.



Fig. 10: Basic scenario frame 600. incPCP, PRMF, DECOLOR fail to detect the static foreground object, though a careful reader can detect a blurry reconstruction of the car in incPCP foreground. However, the SSIM map of all methods are equally good.

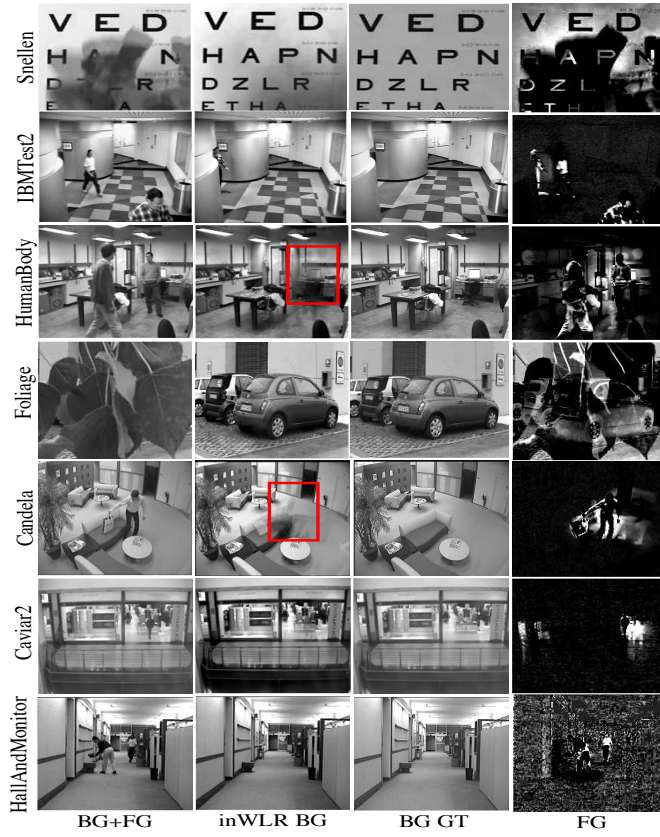


Fig. 11: Qualitative results of inWLR on selected sequences of SBI dataset.

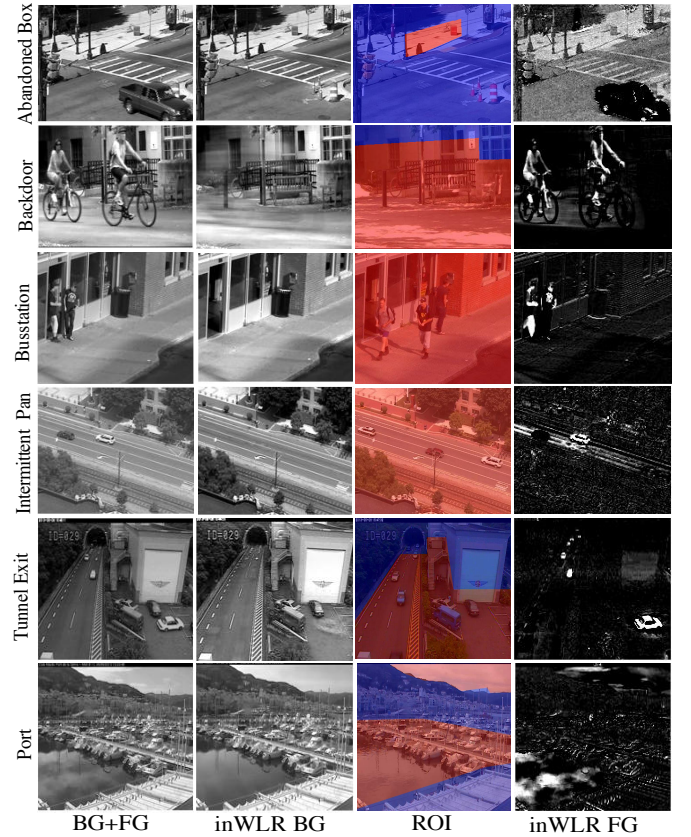


Fig. 12: Qualitative results of inWLR on selected sequences of CDnet 2014 dataset.

it is worth mentioning that inWLR is extraordinarily time efficient compare with the GFL model. We also note that ReProCS is a robust method in removing several background challenges. Although, ReProCS foreground has more false positives than inWLR. This explains why ReProCS suffers in quantitative evaluation (see Figure 8 and 14).

#### D. Performance on SBI dataset

SBI dataset comprises 14 image sequences and they come with 14 background ground truths. Therefore, we can validate

the background recovered by our inWLR algorithm against the background ground truth. For this purpose, we use all four quantitative metrics: area under the ROC curve, SSIM, PSNR, and MSSSIM. To calculate MSSSIM for the SBI dataset, we use a Gaussian window of size  $9 \times 9$  and standard deviation ( $\sigma$ ) 1.5. In Figure 11 we present the qualitative performance of inWLR on 7 different sequences of the SBI dataset and we refer to Table III for quantitative measures. The missing values under the “Area under ROC curve” column of Table III is due to NaN values corresponding the false positive rate



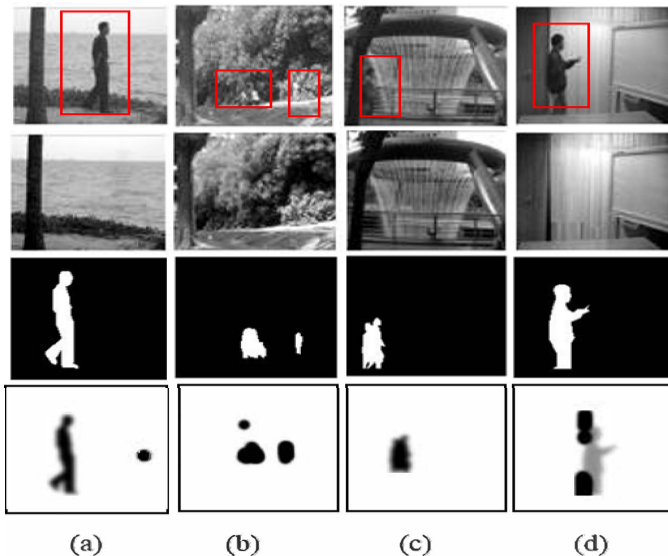


Fig. 13: SSIM index map of: (a) *Water Surface*, (b) *Waving tree*, (c) *Fountain*, and (d) *Meeting Room*. Top to bottom: Original, background estimated by Algorithm 2, ground truth frame (size  $64 \times 80$ ), SSIM index map (size  $54 \times 70$ ) of Algorithm 2. The MSSIM are 0.9851, 0.9082, 0.9940, and 0.9343, respectively.

Video	MSSIM	MSSSIM	PSNR (dB)
Abandoned Box	0.9836	0.9980	58.6752
Backdoor	0.9836	0.9978	60.4562
Busstation	0.9892	0.9987	57.7467
Intermittent Pan	0.9836	0.9978	57.7169
Tunnel Exit	0.9832	0.9977	57.6781
Port	0.9833	0.9978	57.6868
Fountain	0.9886	0.9985	57.6812
Overpass	0.9847	0.998	57.676
Fountain 2	0.9934	0.9992	57.6777
Canoe	0.9858	0.9981	57.678
Fall	0.9853	0.9980	57.7461

TABLE IV: Quantitative performance of inWLR on CDNet 2014 dataset [32].

(FPR) of the predictive analysis. We note that the backgrounds recovered by inWLR sometimes have a minor ghosting effect (see red bounding boxes in Figure 11). However, the average PSNR and MSSSIM of inWLR on 7 sequences of the SBI dataset are 74.3221 and 0.9995, respectively. In contrast, the spatially coherent self-organizing background subtraction (SC-SOBS1) algorithm has the highest average PSNR (35.2723) and MSSSIM (0.9765) on the entire SBI dataset [56], [33]. Recently, the supervised online algorithms, such as, IRLS and Homotopy proposed by Dutta and Richtárik [57], have average MSSSIM on the SBI dataset as 0.9975 and 0.9987, respectively.

#### E. Performance on CDNet2014 dataset

The CDNet 2014 dataset contains a total 11 different video categories. Additionally, each video comes with region of interest (ROI) image which defines the region(s) in a video with foreground movement. The ROI is the red colored part of the frame (with static camera movement). We choose a total 11 sequences from the CDNet 2014 dataset and test the



Fig. 14: *Waving Tree*, frame size  $[120,160]$ . ReProCS and GFL use 220 and 200 pure background frames respectively as training data. The MSSIM for inWLR, GFL, and ReProCS are 0.9592, 0.9996, and 0.5221, respectively. inWLR and GFL recover superior quality background.

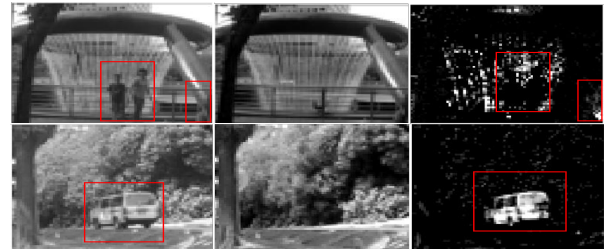


Fig. 15: Top to bottom: *Fountain* with  $p = 5$ , *Campus* with  $p = 6$ . Left to right: Original, inWLR background, and inWLR foreground.

performance of inWLR algorithm both qualitatively and quantitatively. We use SSIM, PSNR, and MSSSIM for quantitative measure. We refer to Figure 12 for the first set of qualitative results. The average PSNR of inWLR on 11 video sequences of CDNet 2014 dataset is 58.0381 dB.

#### F. Dynamic background: a case study

Dynamic background objects are a potential challenge for background modeling algorithms as it is natural for an algorithm to consider them as a part of the foreground. To demonstrate the power of our method on more complex data sets containing dynamic foreground, we perform extensive qualitative and quantitative analysis on the Li data set [29], Wallflower data set [30], and CDNet 2014 dataset [31]. Both our algorithms are capable of detecting the dynamic background objects, although Algorithm 3 is a more natural choice as it deals with the video sequence in an incremental manner. First, in Figure 13, we show the performance of Algorithm 2. The SSIM index map on all four recovered foreground indicates that WLR performs consistently well on the video sequences containing dynamic background. Next, in Figure 14, we compare inWLR against GFL and ReProCS on 60 frames of *Waving Tree* sequence. In Figure 15 we show the performance of inWLR on two data sets with dynamic background and semi-static foreground. Finally, in Figure 16 we present the performance of inWLR on the CDNet 2014 dataset with dynamic background. We also provide the quantitative results in Table IV.

## VI. CONCLUSION

In this paper we proposed two novel algorithms for background estimation. We demonstrated how a properly weighted Frobenius norm can be made robust to the outliers, similarly to the  $\ell_1$  norm in other state-of-the-art background estimation algorithms. Both of our algorithms adaptively determine the

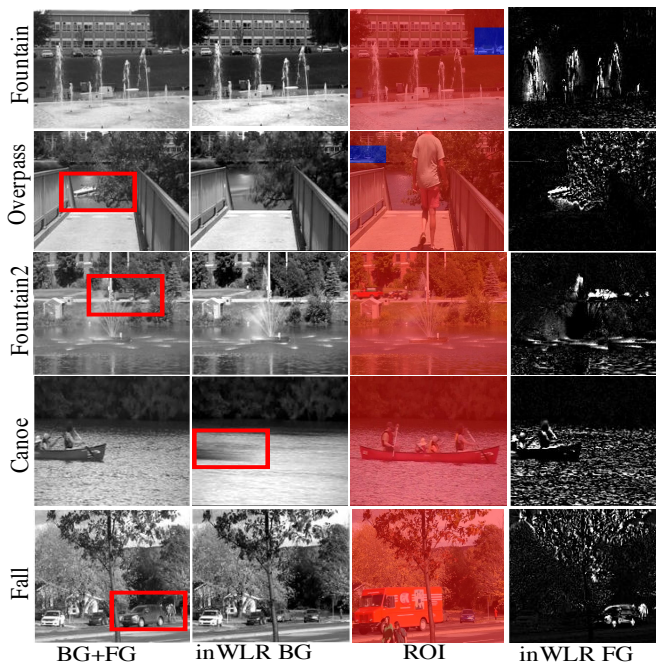


Fig. 16: Qualitative results of inWLR on selected sequences of CDnet 2014 dataset with dynamic background. Only in the *Canoe* sequence, inWLR has some ghosting effect of the foreground object in the background.

background frames without requiring any prior estimate. Furthermore, our batch-incremental algorithm does not require much storage and allows slow changes in the background. Our extensive qualitative and quantitative comparison on real and synthetic video sequences demonstrate the robustness of our algorithms. The batch sizes and the parameters in our incremental algorithm are still empirically selected. Therefore, in future we plan to propose a more robust estimate of the parameters. Like all other algorithms in the literature, we have some limitations as well. Although we are not explicitly required to use pure training background frames, it is mandatory that the video has availability of some pure background frames. Then our algorithm can automatically detect them and use them efficiently in background modeling. Otherwise it will detect the frames with least foreground movements and approximately preserve them as a contender of the background. In the later case the constructed background can be defective.

#### ACKNOWLEDGMENT

This work was supported by the King Abdullah University of Science and Technology (KAUST) Office of Sponsored Research.

#### REFERENCES

- [1] G. H. Golub, A. Hoffman, and G. W. Stewart, "A generalization of the Eckart-Young-Mirsky matrix approximation theorem," *Linear Algebra and its Applications*, vol. 88, no. 89, pp. 317–327, 1987.
- [2] I. T. Jolliffe, "Principal component analysis," 2002, second edition.
- [3] T. Bouwmans, "Traditional and recent approaches in background modeling for foreground detection: An overview," *Computer Science Review*, vol. 11–12, pp. 31 – 66, 2014.
- [4] T. Bouwmans, A. Sobral, S. Javed, S. K. Jung, and E.-H. Zahzah, "Decomposition into low-rank plus additive matrices for background-foreground separation: A review for a comparative evaluation with a large-scale dataset," *Computer Science Review*, 2016.
- [5] A. Sobral and A. Vacavant, "A comprehensive review of background subtraction algorithms evaluated with synthetic and real videos," *Computer Vision and Image Understanding*, vol. 122, pp. 4 – 21, 2014.
- [6] N. Oliver, B. Rosario, and A. Pentland, "A Bayesian computer vision system for modeling human interactions," in *International Conference on Computer Vision Systems*, 1999, pp. 255–272.
- [7] E. J. Candès, X. Li, Y. Ma, and J. Wright, "Robust principal component analysis?" *Journal of the Association for Computing Machinery*, vol. 58, no. 3, pp. 11:1–11:37, 2011.
- [8] J. Wright, Y. Peng, Y. Ma, A. Ganesh, and S. Rao, "Robust principal component analysis: exact recovery of corrupted low-rank matrices by convex optimization," *Proceedings of 22nd Advances in Neural Information Processing systems*, pp. 2080–2088, 2009.
- [9] Z. Lin, M. Chen, and Y. Ma, "The augmented lagrange multiplier method for exact recovery of corrupted low-rank matrices," 2010, arXiv:1009.5055.
- [10] A. Dutta, B. Gong, X. Li, and M. Shah, "Weighted singular value thresholding and its application to background estimation," 2017, arXiv:1707.00133.
- [11] G. Mateos and G. Giannakis, "Robust PCA as bilinear decomposition with outlier-sparsity regularization," *IEEE Transaction on Signal Processing*, vol. 60, no. 10, pp. 5176–5190, 2012.
- [12] N. Wang, T. Yao, J. Wang, and D.-Y. Yeung, "A probabilistic approach to robust matrix factorization," in *Proceedings of 12th European Conference on Computer Vision*, 2012, pp. 126–139.
- [13] J. He, L. Balzano, and A. Szlam, "Incremental gradient on the grassmannian for online foreground and background separation in subsampled video," *IEEE Computer Vision and Pattern Recognition*, pp. 1937–1944, 2012.
- [14] J. Xu, V. K. Ithapu, L. Mukherjee, J. M. Rehg, and V. Singh, "Gosus: Grassmannian online subspace updates with structured sparsity," in *Proceedings of IEEE International Conference on Computer Vision*, 2013, pp. 3376–3383.
- [15] A. Dutta, "Weighted low-rank approximation of matrices:some analytical and numerical aspects," Ph.D. dissertation, University of Central Florida, 2016.
- [16] B. Xin, Y. Tian, Y. Wang, and W. Gao, "Background subtraction via generalized fused Lasso foreground modeling," *IEEE Computer Vision and Pattern Recognition*, pp. 4676–4684, 2015.
- [17] H. Zhao, P. C. Yuen, and J. T. Kwok, "A novel incremental principal component analysis and its application for face recognition," *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, vol. 36, no. 4, pp. 873–886, 2006.
- [18] D. Farcas, C. Marghes, and T. Bouwmans, "Background subtraction via incremental maximum margin criterion: a discriminative subspace approach," *Machine Vision and Applications*, vol. 23, no. 6, pp. 1083–1101, 2012.
- [19] Y. Li, L. Q. Xu, J. Morphet, and R. Jacobs, "An integrated algorithm of incremental and robust pca," in *Proceedings 2003 International Conference on Image Processing*, vol. 1, 2003, pp. I–245–I–248.
- [20] H. Guo, C. Qiu, and N. Vaswani, "An online algorithm for separating sparse and low-dimensional signal sequences from their sum," *IEEE Transactions on Signal Processing*, vol. 62, no. 16, pp. 4284–4297, 2014.
- [21] —, "Practical REPROCS for separating sparse and low-dimensional signal sequences from their sum-part 1," in *IEEE International Conference on Acoustic, Speech and Signal Processing*, 2014, pp. 4161–4165.
- [22] C. Qiu and N. Vaswani, "Support predicted modified-CS for recursive robust principal components pursuit," in *IEEE International Symposium on Information Theory*, 2011, pp. 668–672.
- [23] P. Rodriguez and B. Wohlberg, "Incremental principal component pursuit for video background modeling," *Journal of Mathematical Imaging and Vision*, vol. 55, no. 1, pp. 1–18, 2016.
- [24] —, "A matlab implementation of a fast incremental principal component pursuit algorithm for video background modeling," in *IEEE International Conference on Image Processing*, 2014, pp. 3414–3416.
- [25] —, "Translational and rotational jitter invariant incremental principal component pursuit for video background modeling," in *2015 IEEE International Conference on Image Processing*, 2015, pp. 537–541.
- [26] X. Zhou, C. Yang, and W. Yu, "Moving object detection by detecting contiguous outliers in the low-rank representation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 3, pp. 597–610, 2013.



- [27] A. Dutta and X. Li, "A fast algorithm for a weighted low rank approximation," in *15th IAPR International Conference on Machine Vision Applications (MVA)*, 2017, pp. 93–96.
- [28] S. Brutzer, B. Höferlin, and G. Heidemann, "Evaluation of background subtraction techniques for video surveillance," *IEEE Computer Vision and Pattern Recognition*, pp. 1568–1575, 2012.
- [29] L. Li, W. Huang, I.-H. Gu, and Q. Tian, "Statistical modeling of complex backgrounds for foreground object detection," *IEEE Transactions on Image Processing*, vol. 13, no. 11, pp. 1459–1472, 2004.
- [30] K. Toyama, J. Krumm, B. Brumitt, and B. Meyers, "Wallflower: Principles and practice of background maintenance," *Seventh International Conference on Computer Vision*, pp. 255–261, 1999.
- [31] Y. Wang, P. M. Jodoin, F. Porikli, J. Konrad, Y. Benezeth, and P. Ishwar, "Cdnnet 2014: An expanded change detection benchmark dataset," in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2014, pp. 393–400.
- [32] L. Maddalena and A. Petrosino, "Towards benchmarking scene background initialization," in *New Trends in Image Analysis and Processing – ICIAP 2015 Workshops*, 2015, pp. 469–476.
- [33] [Http://sbmi2015.na.icar.cnr.it/SBIdataset.html](http://sbmi2015.na.icar.cnr.it/SBIdataset.html).
- [34] A. Dutta and X. Li, "On a problem of weighted low-rank approximation of matrices," *SIAM Journal on Matrix Analysis and Applications*, vol. 38, no. 2, pp. 530–553, 2017.
- [35] K. Usevich and I. Markovsky, "Variable projection methods for affinely structured low-rank approximation in weighted 2-norms," *Journal of Computational and Applied Mathematics*, vol. 272, pp. 430–448, 2014.
- [36] I. Markovsky, J. C. Willems, B. D. Moor, and S. V. Huffel, "Exact and approximate modeling of linear systems: a behavioral approach," 2006, SIAM.
- [37] I. Markovsky, "Low-rank approximation: algorithms, implementation, applications," 2012, Springer.
- [38] K. Usevich and I. Markovsky, "Optimization on a grassmann manifold with application to system identification," *Automatica*, vol. 50, no. 6, pp. 1656–1662, 2014.
- [39] T. Okatani and K. Deguchi, "On the Wiberg algorithm for matrix factorization in the presence of missing components," *International Journal of Computer Vision*, vol. 72, no. 3, pp. 329–337, 2007.
- [40] T. Wiberg, "Computation of principal components when data are missing," in *Proceedings of the Second Symposium of Computational Statistics*, pp. 229–236, 1976.
- [41] N. Srebro and T. Jaakkola, "Weighted low-rank approximations," *20th International Conference on Machine Learning*, pp. 720–727, 2003.
- [42] N. Srebro, J. D. M. Rennie, and T. S. Jaakkola, "Maximum-margin matrix factorization," in *Proceedings of Advances in Neural Information Processing Systems*, vol. 18, pp. 1329–1336, 2005.
- [43] A. M. Buchanan and A. W. Fitzgibbon, "Damped Newton algorithms for matrix factorization with missing data," in *Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 2, pp. 316–322, 2005.
- [44] J. H. Manton, R. Mehony, and Y. Hua, "The geometry of weighted low-rank approximations," *IEEE Transactions on Signal Processing*, vol. 51, no. 2, pp. 500–514, 2003.
- [45] W. S. Lu, S. C. Pei, and P. H. Wang, "Weighted low-rank approximation of general complex matrices and its application in the design of 2-d digital filters," *IEEE Transactions on Circuits and Systems I: Fundamental Theory and Applications*, vol. 44, no. 7, pp. 650–655, 1997.
- [46] D. Shpak, "A weighted-least-squares matrix decomposition with application to the design of 2-d digital filters," *Proceedings of IEEE 33rd Midwest Symposium on Circuits and Systems*, pp. 1070–1073, 1990.
- [47] T. Hastie, R. Mazumder, J. Lee, and R. Zadeh, "Matrix completion and low-rank svd via fast alternating least squares," *Journal of Machine Learning Research*, vol. 16, pp. 1–122, 2011.
- [48] J. Hansohm, "Some properties of the normed alternating least squares (als) algorithm," *Optimization*, vol. 19, pp. 1–122, 2011.
- [49] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Foundations and Trends in Machine Learning*, vol. 3, no. 1, pp. 1–122, 2011.
- [50] J. F. Cai, E. J. Candès, and Z. Shen, "A singular value thresholding algorithm for matrix completion," *SIAM Journal on Optimization*, vol. 20, no. 4, pp. 1956–1982, 2010.
- [51] A. Dutta and X. Li, "Weighted low rank approximation for background estimation problems," in *The IEEE International Conference on Computer Vision Workshops (ICCVW)*, 2017, pp. 1853–1861.
- [52] A. Dutta, X. Li, and P. Richtárik, "A batch-incremental video background estimation model using weighted low-rank approximation of matrices," in *The IEEE International Conference on Computer Vision Workshops (ICCVW)*, 2017, pp. 1835–1843.
- [53] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE Transaction on Image Processing*, vol. 13, no. 4, pp. 600–612, 2004.
- [54] Z. Wang, E. P. Simoncelli, and A. C. Bovik, "Multi-scale structural similarity for image quality assessment," in *37th IEEE Asilomar Conference on Signals, Systems, and Computers*, 2003, pp. 1398–1402.
- [55] T. Bouwmans, L. Maddalena, and A. Petrosino, "Scene background initialization: A taxonomy," *Pattern Recognition Letters*, vol. 96, pp. 3–11, 2017.
- [56] L. Maddalena and A. Petrosino, "The SOBS algorithm: What are the limits?" in *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2012, pp. 21–26.
- [57] A. Dutta and P. Richtárik, "Online and batch supervised background estimation via L1 regression," 2017, arXiv:1712.02249.



Abdullah University of Science and Technology (KAUST), Saudi Arabia. His research interests include weighted and structured low-rank approximation of matrices, convex, nonlinear, and stochastic optimization, numerical analysis, linear algebra, and deep learning. In addition, he works on the applications of image and video analysis in computer vision and machine learning.



**Xin Li** Dr. Xin Li is Professor and Chair of Mathematics at the University of Central Florida. He received his B.S. and M.S. degrees from Zhejiang University in 1983 and 1986, respectively, and his PhD in Mathematics from the University of South Florida in 1989. Dr. Li's research interests include Approximation Theory and its applications in scientific computing, machine learning, and computer vision.



**Peter Richtárik** is an Associate Professor of Computer Science and Mathematics at KAUST and an Associate Professor of Mathematics at the University of Edinburgh. He is an EPSRC Fellow in Mathematical Sciences, Fellow of the Alan Turing Institute, and is affiliated with the Visual Computing Center and the Extreme Computing Research Center at KAUST. Dr. Richtárik received his PhD from Cornell University in 2007, and then worked as a Postdoctoral Fellow in Louvain, Belgium, before joining Edinburgh in 2009, and KAUST in 2017.

Dr. Richtárik's research interests lie at the intersection of mathematics, computer science, machine learning, optimization, numerical linear algebra, high performance computing, and applied probability. Through his recent work on randomized decomposition algorithms (such as randomized coordinate descent methods, stochastic gradient descent methods and their numerous extensions, improvements and variants), he has contributed to the foundations of the emerging field of big data optimization, randomized numerical linear algebra, and stochastic methods for empirical risk minimization. Several of his papers attracted international awards, including the SIAM SIGEST Best Paper Award and the IMA Leslie Fox Prize (2nd prize, three times). He is the founder and organizer of the Optimization and Big Data workshop series.